Weekly Report

1 Done

1.1 Funding Summarization

Collect paper and patents and start reading.

1.2 Project

- Introduce my idea with Chris.
- Read related papers and write related work. The paper can be checked at: https://www.overleaf.com/read/ynbpgxjsngfy

Inference Attack against Encrypted Range Queries on Outsourced Databases

"a general attack on Precise Query Protocols (PQPs, without incurring any false positives) based on access pattern disclosure in the context of secure range queries"

"the attacker has reasonable amount background knowledge"

Inference Attacks on Property-Preserving Encrypted Databases

"The most well-known example of an inference attack is frequency analysis which is used to break classical ciphers. Another example is the query-recovery attack of Islam et al. against searchable symmetric encryption (SSE) schemes"

- "Individual attacks: recover a row in the database.
- Aggregate attacks: recover statistical information about the entire database."

Access Pattern disclosure on Searchable Encryption: Ramification, Attack and Mitigation

Sensitive information may leak when "searching over remote encrypted data". "It is assumed that the server does not have access to the trapdoor generation function, and therefore, can not ascertain the keyword searched for. However, it is imperative to hide the corresponding keyword of a given query from an adversary." Infer sensitive information based on the joint probability distribution for two or more variables.

A Novel Attack Graph Posterior Inference Model Based on Bayesian Network

Zhang and Song proposed a graph-based posterior inference model based on likelihood weighting. Considering various possibilities of the event sequences, their graph model includes circles and bi-directed edges.

Preventing Private Information Inference Attacks on Social Networks

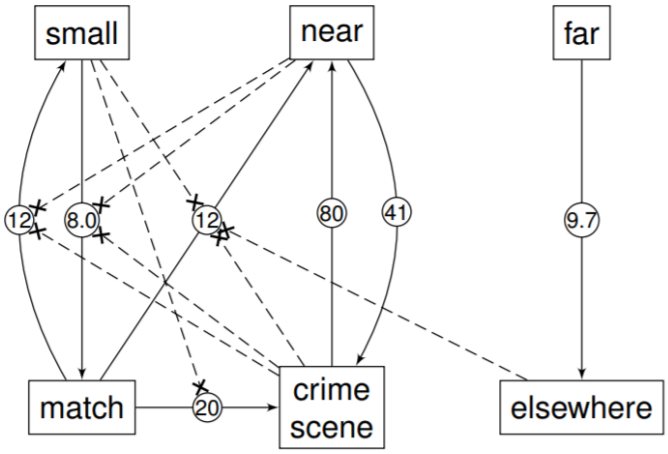
Modeled data; Naïve Bayes; Social network data with attributed nodes; Weighting friendships.

Toward inference attacks for k-anonymity

"Traditional k-anonymity and its extended models are more vulnerable in face of attackers that can get background knowledge from big data. … Lacking the background knowledge about attackers and its attack behaviors is the main reason that those models still not effectively prevent the

privacy reasoning destruction in k-anonymous data sets."
They presented a privacy attack graph. "Privacy attack graph for k-anonymity implicitly describes anonymous attacker to exploit the available data sets and background knowledge, reasoning to get all the inference attack paths of user privacy step by step."

Inference and Attack in Bayesian Networks
"A drawback of the use of Bayesian networks is that a BN's meaning is not very intuitive and therefore hard to understand. Supporters of this approach often argue that the structure of the graph is an intuitive representation of the relation between the variables, but in reality the meaning of the graph is often not easily explained. This is due to the fact that edges, and in particular the direction of the edges, suggest causality, but in fact have no intuitive interpretation in Bayesian networks."
"The inference rules and their strength extracted from the example BN. Every inference rule is displayed as an arrow with a small circle halfway that states the strength. undercutters are displayed as a dashed, cross-tipped arrow pointing to the circle of the inference. Only inferences with a strength greater than five are shown to prevent visual clutter."



Kl-anonymity
To resist identity inference, Li et al. presented a flexible algorithm that combined k-anonymity and l-diversity by employing two parameters--k as individual number and l as diversity level, simultaneously.

Disclose more and risk less
Chen et al. convert the trade-off between privacy and utility to a knapsack problem by regarding attribute values as "items".
The "weight" is the quantified privacy disclosure risk calculated based on Naïve Bayes model, and the "profit" is related to uniqueness and commonness.

Minimax Filter
"This work presents a new learning-based mechanism for preventing inference attacks on continuous and high-dimensional data. In this mechanism, a filter transforms continuous and high-dimensional raw features to dimensionality-reduced representations of data. After

filtering, information on target tasks remains but information on identifying or sensitive attributes is removed which makes it difficult for an adversary to accurately infer such attributes from the released filtered output."

Private and Public

"We propose a practical methodology to protect a user's private data, when he wishes to publicly release data that is correlated with his private data, to get some utility. Our approach relies on a general statistical inference framework that captures the privacy threat under inference attacks, given utility constraints. Under this framework, data is distorted before it is released, according to a probabilistic privacy mapping. This mapping is obtained by solving a convex optimization problem, which minimizes information leakage under a distortion constraint."

Mapping based on mutual information.

1.3 Discussion with Prof. Deng

He talked about two directions about privacy preservation:

- How to process data before sharing data with peer company to provide better service? (Actually, this issue has been studied by other researchers. Eg. Federated Transfer Learning)
- Verify if the privacy is preserved by implanting fake identities. Those identities should have not only uniqueness for easy to identify from public data, but also commonness to avoid from detecting by data owners.

2 Work Hours

Weekdays: 10:30-17:10 at lab and 19:30-21:30 at home.
Monday: 17:10-18:00 for group meeting.
Weekend: 16:00-18:00 and 19:30-21:00 at home.